TRANSLATION OF THE AMENDED PAGE

2/PRTS

09/674962

528 Rec'd PCT/PTO 08 NOV 2000

Novel peptide fragments for purifying proteins

The invention relates to novel peptide fragments, fusion proteins comprising the peptide fragments, processes for preparing them, and the use of the peptide fragments. The invention also relates to a process for purifying fusion proteins and a method for detecting proteins.

The invention further relates to nucleic acids which code for the peptide fragments or for the fusion proteins, and vectors which comprise the nucleic acids.

It has become possible through modern molecular biology to prepare almost any proteins, peptides or their derivatives in virtually unlimited quantities. In this connection, purification of proteins frequently proves to be the limiting and not uncommonly inefficient and thus eventually cost-determining factor.

This is why a number of techniques for purifying proteins have been developed. The techniques normally used to purify proteins from cell supernatants, crude cell extracts or cells are, for example, salt precipitation or precipitation with organic solvents, ultrafiltration, dialysis, gel electrophoresis, isoelectric focusing, chromatofocusing, ion exchange or gel chromatography, hydrophobic chromatography, immunoprecipitation or IMAC (= immobilized metal affinity chromatography). Single methods can be used for the purification, but a combination of different techniques is usually necessary. Conventional purification methods are reviewed, for example, in the textbooks Protein Purification Process Engineering (Ed. R.G. Harrison, 1994, Marcel Dekker, Inc., New York, pages 209 - 316, ISBN 0-8247-9009-X), Protein Purification (Ed. R.K. Scopes, 1994, Springer Verlag New York, chapter 4 - 7, ISBN 0-387-94072-3) and Methods for Protein Analysis (Ed. R.A. Copeland, 1994 Chapman & Hall, page 59 - 112, ISBN 0-412-03741-6). It is important for the purification of proteins that the purification takes place under conditions which are as mild as possible, and is as selective as possible and as quick as possible. Moreover the protein losses should be kept as small as possible. Many of the protein purification methods are insufficiently selective, suitable for only small quantities of protein and/or very costly.

The method for purifying proteins by IMAC (= immobilized metal affinity chromatography) described by Porath et al. (Nature, Vol.258, 1975: 598 - 599) is a good compromise between the stated requirements to be met by optimal purification. However, this

method still has some disadvantages. Thus, for example, not all metal ions bind equally well to the support material so that some of the ions are washed out and thus contaminate the required product. Many proteins do not bind at all to the chromatography material and thus cannot be purified, or bind too weakly so that they are eluted even during the necessary steps for washing the column material. This results in undesired losses of product. Since the selectivity is usually inadequate for a one-step purification, in contrast to purifications by biospecific affinity purification methods such as, for example, purification via antibodies, further purification steps are necessary.

In order to be able to purify a wider range of proteins using this method, various so-called tags have been developed, such as polyhistidine, His-Trp, His-Tyr or $(His\text{-}Asp)_n$ (see Sporeno et al., J. Biol. Chem. 269 (15), 1994: 10991 - 10995, Le Grice et al., Eur. J. Biochem., 187 (2), 1990: 307 - 314, Reece et al., Gene, 126 (1), 1993: 105 - 107, De Vos et al., Nucl. Acid. Res., 22 (7), 1994: 1161 - 1166, Feng et al., J. Biol. Chem. 269 (3), 1994: 2342 - 2348, Hochuli et al., Biotechnology, 1988: 1321 - 1325, Patwardhan et al., J. Chromatography A, 787, 1997: 91 - 100, Hutchens et al., J. Chromatography, 604, 1992: 133 - 141). These tags are linked to the protein to be purified by means of molecular biology at the nucleic acid level. It has been possible through these tags to improve protein purification further in some areas. However, even this method still has some disadvantages. It is still not possible to predict reliably whether a protein can be purified by this method (see Immobilized Metal Ion Affinity Chromatography, L. Kågedal, page 227 - 251 in Protein Purification, Eds. J.C. Janson, L. Rydén, 1989, VCH Publishers, Inc., New York, ISBN 0-89573-122-3), which means that this method is not applicable to every protein either. Once again, metal ions may be washed out or proteins may bind so weakly that they are partly lost during the washing steps. The selectivity is still unsatisfactory too. In addition, the capacity of the column material for loading with the proteins to be purified is still in some cases too low, so that a large amount of column material must be used for a purification. The protein yield is also still inadequate. This leads to unnecessary costs.

Volz et al. have described the purification of ATPases from Helicobacter pylori without using an additional His-tag sequence. These ATPases contain natural metal binding sites which make purification by IMAC possible. Our own studies have shown that these binding sites display a binding affinity which is too low for efficient purification of all desired proteins.

It is an object of the present invention to provide further tags for protein purification by IMAC which do not have the abovementioned disadvantages and thus make it possible, for example, to use the tags more widely and/or load the column material with a greater density, and which show a higher selectivity and thus simplify the purification.

We have found that this object is achieved by the peptide fragments according to the invention having the general sequence

His-X^1-His-X^2-X^3-X^4-Cys-X^5-X^6-Cys,

where the variables X^1 to X^6 in the sequence have the following meanings:

X^1 = an amino acid selected from the group of Ala, Val, Phe, Ser, Met, Trp, Tyr, Asn, Asp or Lys and the variables X^2 to X^6 an amino acid selected from the group of Gly, Ala, Val, Leu, Ile, Phe, Pro, Ser, Thr, Cys, Met, Trp, Tyr, Asn, Gln, Asp, Glu, Lys, Arg, His or

X^2 = an amino acid selected from the group of Val, Ile, Phe, Pro, Trp, Tyr, Gln, Glu or Arg and the variables X^1, X^3 to X^6 an amino acid selected from the group of Gly, Ala, Val, Leu, Ile, Phe, Pro, Ser, Thr, Cys, Met, Trp, Tyr, Asn, Gln, Asp, Glu, Lys, Arg, His or

X^3 = an amino acid selected from the group of Gly, Ile, Thr, Met, Trp, Tyr, Asn, Gln, Asp, Glu, Lys, Arg, His and the variables X^1, X^2, X^4 to X^6 an amino acid selected from the group of Gly, Ala, Val, Leu, Ile, Phe, Pro, Ser, Thr, Cys, Met, Trp, Tyr, Asn, Gln, Asp, Glu, Lys, Arg, His or

X^4 = an amino acid selected from the group of Val, Phe, Pro, Cys, Met, Trp, Asn, Glu, Arg or His and the variables X^1 to X^3, X^5, X^6 an amino acid selected from the group of Gly, Ala, Val, Leu, Ile, Phe, Pro, Ser, Thr, Cys, Met, Trp, Tyr, Asn, Gln, Asp, Glu, Lys, Arg, His or

X^5 = an amino acid selected from the group of Gly, Ser, Cys, Met, Trp, Asn, Glu, Lys or Arg and the variables X^1 to X^4, X^6 an amino acid selected from the group of Gly, Ala, Val, Leu, Ile, Phe, Pro, Ser, Thr, Cys, Met, Trp, Tyr, Asn, Gln, Asp, Glu, Lys, Arg, His or

X^6 = an amino acid selected from the group of Phe, Pro, Ser, Cys, Trp, Tyr or Gln and the variables X^1 to X^5 an amino acid selected from the group of Gly, Ala, Val, Leu, Ile, Phe, Pro, Ser, Thr, Cys, Met, Trp, Tyr, Asn, Gln, Asp, Glu, Lys, Arg, His and

where at least one of the variabes X^1 to X^6 in the sequence is, independently of one another, Gln or Asn.

~~The general sequence His-X^1-His-X^2-X^3-X^4-Cys-X^5-X^6-Cys corresponds to SEQ ID No: 1 where X^1 corresponds to the amino acids designated Xaa in position 2 in SEQ ID NO: 1, and X^2 corresponds to Xaa in position 4, X^3 corresponds to Xaa in position 5, X^4 corresponds to Xaa in position 6, X^5 corresponds to Xaa in position 8 and X^6 corresponds to Xaa in position 9. The amino acids mentioned above for X^1 to X^6 may represent the corresponding amino acids designated Xaa in SEQ ID NO: 1.~~

It is advantageous for at least one of the variables X^1 to X^6 in the sequence additionally to be, independently of one another, Lys or Arg. Further advantageous amino acids present in the variables X^1 to X^6 are Glu, Lys, Arg, Tyr, Cys, Lys, His, Asp or Met. The amino acids Cys, Glu, Lys, Tyr or Arg are preferably present, particularly preferably Cys. These amino acids contribute to advantageous binding of the peptide fragments to the immobilized metal ions. In addition, it is advantageous for not more than four, preferably three, histidine residues to be present consecutively in the sequence.

The variables X^1 to X^6 in the sequence have the further advantageous and preferred meanings, independently of one another:

X^1 = an amino acid selected from the group of Ala, Val, Phe, Ser, Met, Trp, Tyr, Asn, Asp or Lys, particularly preferably Phe, Ser, Asn, Asp or Lys, very particularly preferably Asn;

X^2 = an amino acid selected from the group of Val, Ile, Phe, Pro, Trp, Tyr, Gln, Glu or Arg, particularly preferably Val, Ile, Phe, Pro, Gln, Glu or Arg, very particularly preferably Gln, Glu or Arg;

X^3 = an amino acid selected from the group of Gly, Ile, Thr, Met, Trp, Tyr, Asn, Gln, Asp, Glu, Lys, Arg or His, particularly preferably Gly, Ile, Thr, Met, Trp, Tyr, Asn, Asp, Glu, Arg or His, very particularly preferably Gly, Thr or Tyr;

X^4 = an amino acid selected from the group of Val, Phe, Pro, Cys, Met, Trp, Asn, Glu, Arg or His, particularly preferably Val, Phe, Cys, Met, Trp, Asn, Arg or His, very particularly preferably Asn or Arg;

X^5 = an amino acid selected from the group of Gly, Ser, Cys, Met, Trp, Asn, Glu, Lys or Arg, particularly preferably Gly, Ser, Cys, Met, Asn, Glu, Lys or Arg, very particularly preferably Gly or Lys;

X^6 = an amino acid selected from the group of Phe, Pro, Ser, Cys, Trp, Tyr or Gln, particularly preferably Phe, Ser, Cys or Tyr, very particularly preferably Cys.

Sub C2

The variables X^1 to X^6 in the sequence His-X^1-His-X^2-X^3-X^4-Cys-X^5-X^6-Cys may have the various preferred meanings independently of one another, with individual variables up to a maximum of five of the variables being an amino acid selected from the group of Gly, Ala, Val, Leu, Ile, Phe, Pro, Ser, Thr, Cys, Met, Trp, Tyr, Asn, Gln, Asp, Glu, Lys, Arg, His.

Particularly preferred peptide fragments are fragments having the sequences

His-Gln-His-Glu-Gly-Arg-Cys-Lys-Glu-Cys

His-Asn-His-Arg-Tyr-Gly-Cys-Gly-Cys-Cys

His-Arg-His-Gly-Thr-Asn-Cys-Leu-Lys-Cys

His-Ile-His-Gln-Ser-Asn-Cys-Gln-Val-Cys.

The stated sequences correspond in each case to the sequences SEQ ID NO: 2, SEQ ID NO: 3, SEQ ID NO: 4 and SEQ ID NO: 5.

The abovementioned protein fragment sequence is encoded by the nucleic acid fragments according to the invention. Account must be taken of the degenerate genetic code in this connection. The nucleic acid fragments according to the invention may in principle be present in any suitable nucleic acids. The nucleic acid fragments are advantageously inserted into vectors in such a way that it is possible to prepare composite nucleic acid sequences (= gene constructs) which code for the fusion proteins according to the invention. These gene constructs can, for expression, advantageously be accommodated in a suitable host organism which makes optimal expression of the fusion protein possible. Suitable vectors are well known to the skilled worker

and can be found, for example in the book Cloning Vectors (Eds. Pouwels P. H. et al. Elsevier, Amsterdam-New York-Oxford, 1985, ISBN 0 444 904018). Apart from plasmids, vectors mean all other vectors known to the skilled worker, such as, for example, phages, viruses, transposons, IS elements, plasmids, cosmids, linear or circular DNA. These vectors may undergo autonomous replication or chromosomal replication in the host organisms.

The nucleic acid sequences according to the invention mean sequences which have been functionally linked to one or more regulatory signals, advantageously to increase gene expression. These sequences may be 3′ and/or 5′ terminal regulatory sequences to enhance expression and are selected for optimal expression depending on the selected host organism and gene. These regulatory sequences are, for example, sequences to which inducers or repressors bind and thus regulate the expression of the nucleic acids. The gene construct may additionally advantageously contain one or more so-called enhancer sequences functionally linked to the promoter, and these make increased expression of the nucleic acid sequence possible. This may take place, for example, by an approved interaction between RNA polymerase and DNA. It is also possible to insert additional advantageous sequences at the 3′ end of the DNA sequences, such as other regulatory elements or terminators. The nucleic acid fragments according to the invention are advantageously inserted into the vector in such a way that they form the N-terminal region of the fusion protein. However, they can also be located at the C terminus, or else be located within the protein, but in this case the function of the protein must not be affected, and cutting out of the fusion protein is no longer possible.

The regulatory sequences are intended to make specific expression of the genes and protein expression possible. This may mean, for example, depending on the host organism that the gene is expressed or overexpressed only after induction, or that it is immediately expressed and/or overexpressed.

Advantageous regulatory sequences are present, for example, in promoters such as the cos, tac, trp, tet, trp-tet, lpp, lac, lpp-lac, $lacI^q$, T7, T5, T3, gal, trc, ara, SP6, λ-P_R promoter or in the λ-P_L promoter, which are advantageosly used in Gram-negative bacteria. Further advantageous regulatory sequences are, for example, present in the Gram-positive promoters amy and SPO2, in the yeast or fungus promoters ADC1, MFα , AC, P-60, CYC1, GAPDH, TEF, rp28, ADH or in the plant promoters CaMV/35S, SSU, OCS, lib4, usp, STLS1, B33, nos or in the ubiquitin or phaseolin promoter. Also advantageous in this connection are the promoters of pyruvate decarboxylase and of methanol oxidase from,

for example, Hansenula. It is also possible to use artificial promoters for the regulation.

The regulatory sequences or factors may moreover preferably have a beneficial effect on expression of the introduced gene, and thus increase it. Thus, enhancement of the regulatory elements can advantageously take place at the level of transcription by using strong transcription signals such as said promoters and/or enhancers. However, enhancement of translation is also possible by, for example, improving the stability of the mRNA.

The nucleic acid sequences according to the invention advantageously also contain signals which make it possible for the proteins to be secreted into the medium or into cell compartments. Examples of sequences of this type which may be mentioned are the typical signal sequences such as, for example the signal sequence of ompA (E. coli membrane protein). It is additionally possible for other advantageous sequences to be present, such as the seqeunce of the α factor, or for YACs (= yeast artificial chromosomes) to be used.

The gene constructs (= nucleic acid sequences) according to the invention advantageously additionally contain sequences which make it possible to eliminate the protein fragments having the abovementioned sequence according to the invention from the N or C terminus of the fusion protein, preferably from the N terminus. These sequences code, for example, for cleavage sites for a wide variety of proteases such as, for example, for factor Xa, enterokinase, human renin, carboxypeptidase A, thrombin, trypsin, dipeptidyl peptidases, papain, plasmin, pepsin or other proteases. Preferred cleavage sites are for factor Xa, human renin, dipeptidyl peptidases, carboxypeptidase A or enterokinase, because these enzymes have a high specificity and thus unwanted digestion of the protein to be purified can be avoided. If other proteases are used, care must be taken that no cleavage sites are inside the protein to be purified. The protein fragment can also be deleted by cleavage with cyanogen bromide [e.g. 2-(2-nitrophenylsulfenyl)-3-bromo-3′-methylindolinium, hydroxylamine etc.] in the presence of formic acid. However, in this case refolding of the protein is usually necessary, which results in this method being less preferred. Elimination of the protein fragment by exoprotease digestion (under kinetic control) is also possible. However, this usually results in product mixtures. The cleavage sites preferably used are those making it possible to delete the protein fragments without leaving residues of protein fragments in the protein to be purified. If the protein fragments according to the invention can be tolerated in the fusion protein

without loss of function and without other disadvantages, a specific site for detaching the protein fragment can be dispensed with.

Suitable vectors are in principle all vectors which make expression in pro- or eukaryotic cells possible. It is possible in this connection to use vectors which replicate in only one genus or those which replicate in several genera (called shuttle vectors). Examples of advantageous vectors are plasmids such as the E.coli plasmids pEGFP, pLG338, pACYC184, pBR322, pUC18, pUC19, pKC30, pRep4, pHS1, pHS2, pPLc236, pMBL24, pLG200, pUR290, pIN-III^{113}-B1, λgt11 or pBdCI, preferably pEGFP, in Streptomyces pIJ101, pIJ364, pIJ702 or pIJ361, in Bacillus pUB110, pC194 or pBD214, in Corynebacterium pSA77 or pAJ667, in fungi pALS1, pIL2 or pBB116, in yeasts 2μm, pAG-1, YEp6, YEp13 or pEMBLYe23 or in plants pLGV23, pGHlac^{+}, pBIN19, pAC2004 or pDH51. Said plasmids represent a small selection of the possible plasmids. Further plasmids are well known to the skilled worker and can be found, for example, in the abovementioned book Cloning Vectors (Eds. Pouwels P. H. et al. Elsevier, Amsterdam-New York-Oxford, 1985 , ISBN 0 444 904018).

In another embodiment of the vector, the nucleic acid sequence according to the invention can also advantageously be introduced in the form of a linear DNA into the microorganisms and be integrated by heterologous or homologous recombination into the genome of the host organism. This linear DNA can consist of a linearized vector such as a plasmid or only of the nucleic acid, i.e. the nucleic acid fragment and the gene for the protein (= fusion protein gene), and, where appropriate, other regulatory sequences.

Host organisms suitable for the gene construct according to the invention are in principle all prokaryotic or eukaryotic organisms. Host organisms which are advantageously used are microorganisms such as Gram-positive or Gram-negative bacteria, archaebacteria, fungi, yeasts, animal or plant cells such as Drosophila, specifically D.melanogaster, mouse, zebra fish or tobacco. Preferably used are Gram-positive or Gram-negative bacteria, fungi or yeasts, particularly preferably the genera Escherichia, Bacillus, Streptomyces, Aspergillus or Saccharomyces, very particularly preferably E. coli.

Particular preference is given to the following combinations of vector and host organisms such as Escherichia coli and its plasmids and phages and the promoters belonging thereto, and Bacillus and its plasmids and promoters, Streptomyces and its plasmids and promoters, Aspergillus and its plasmids and

promoters or Saccharomyces and its plasmids and promoters.

The fusion proteins according to the invention can be prepared as described above in a process wherein the nucleic acid fragments according to the invention, which code for the protein fragments having the abovementioned sequence, are fused to a gene which codes for the proteins to be purified and, where appropriate, other advantageous sequences such as promoter and/or enhancer sequences, cleavage sites for proteases etc. If necessary for this purpose, a suitable restriction enzyme cleavage site is introduced between the nucleic acid fragment and the gene of the protein to be purified, and this construct is inserted via suitable cleavage sites into a vector. Methods of this type are known to the skilled worker and can be found for example, in the textbooks by Sambrook, J. et al. (1989) Molecular cloning: A laboratory manual, Cold Spring Harbor Laboratory Press, by F.M. Ausubel et al. (1994) Current protocols in molecular biology, John Wiley and Sons or D.M. Glover et al., DNA Cloning Vol.1, (1995), IRL Press (ISBN 019-963476-9). Further advantageous vectors are the Pichia pastoris vectors pPic and pGap. This yeast is also a suitable host organism for the protein expression.

The protein fragments according to the invention are suitable for preparing fusion proteins which can be purified easily, at low cost and efficiently with the aid of the protein fragments. The protein fragments and fusion proteins according to the invention can thus be purified advantageously, very selectively and in good yields and high purity. The protein fragments according to the invention, and thus the fusion proteins prepared from them are advantageously distinguished by binding to the immobilized metal ions at least 1.5 times more strongly than the Helicobacter pylori ATPase-439 sequence.

All proteins are suitable in principle for preparing the fusion proteins. The proteins preferably used are those having a biological effect in humans, animals or plants or those of interest for organic synthesis. Examples thereof are proteins such as enzymes, hormones, or storage or binding or transport proteins. Examples which may be mentioned are proteins such as hydrolases such as lipases, esterases, amidases, nitrilases, proteases, mediators such as cytokines e.g. lymphokines such as MIF, MAF, TNF, interleukins such as interleukin 1, interferons such as γ-interferon, tPA, hormones such as proteohormones, glycohormones, oligo- or polypeptide hormones such as vasopressin, endorphins, endostatin, angiostatin, growth factors, erythropoietin, transcription factors, integrins such as GPIIb/IIIa or $\alpha_v\beta$III, receptors such as the various glutamate receptors,

angiogenesis factors such as angiotensin.

The process according to the invention for purifying fusion proteins makes it possible, for example, to purify proteins from natural sources such as plant or animal extracts, plant or animal cell lysates, from culture media, fermentation broths or from synthesis broths, to mention only a few by way of example.

The process according to the invention comprises the following reaction steps:

a) bringing liquids which contain the fusion protein into contact with immobilized metal ions in such a way that an affinity linkage can form between the metal ions and the fusion protein,

b) removing unbound substances present in the liquid,

c) eluting the bound fusion protein by abolishing the affinity linkage by changing the liquid medium and

d) collecting the purified fusion protein.

The fusion protein is advantageously expressed in a suitable host organism (see above) before the purification in order to increase the yield of fusion protein. The host organism is cultured in a suitable synthetic or complex medium which contains a carbon source, a nitrogen source and, where appropriate, inorganic salts, vitamins and trace elements, at a suitable temperature and with suitable aeration.

Depending on whether the fusion protein is excreted from the cells or not, the cells are first disrupted, and the cells or cell detritus are advantageously removed. The methods used for cell disruption are those known to the skilled worker, such as ultrasound, French press, enzyme digestion, osmotic shock and many others. The cells or cell detritus can be removed, for example, by centrifugation or filtration. However, removal of the cells or cell detritus is not absolutely necessary.

The liquid containing the fusion proteins is subsequently brought into contact with the immobilized metal ions so that an affinity linkage between the fusion protein and the metal ions can form. The binding takes place at pH values greater than 7, advantageously for example at pH 7.0 to 9.0, preferably between pH 7.5 and 8.0. Advantageous buffers are single buffers or buffer mixtures such as, for example, 50 to 1000 mM buffers such as

50 mM Tris/HCl pH 8.0 + 150 mM NaCl, 100 mM sodium acetate pH 7.7 + 500 mM NaCl, 20 mM sodium phosphate pH 7.7 + 500 mM NaCl or 50 mM Tris/HCl pH 8.0 + 150 mM NH_4Cl. These buffers make it possible to load the fusion proteins onto the immobilized metal ions. In the simplest and particularly advantageous case, the fusion proteins are brought into contact with the immobilized metal ions directly in the buffer used for the disruption or in the incubation medium. It is advantageous for the liquids and the immobilized metal ions to be brought into contact with one another in a conventional chromatography column. This facilitates the removal of unbound substances, for example proteins, by washing the column with a suitable buffer. Suitable buffers are buffers which do not impair the binding of the protein fragments according to the invention or of the fusion proteins to the metal ions, and are able to remove impurities. Buffers of this type preferably have a pH above pH 7, advantageously a pH between pH 7.0 and 9.0, preferably between pH 7.5 and 8.0. It is also possible to purify batch mixtures, in which the immobilized metal ions are placed in a vessel and then the liquids are added or vice versa, in this way. The buffers which have been mentioned can be used for these batch mixtures. The mixtures are advantageously centrifuged or filtered between the individual washing steps.

Support materials suitable in principle for immobilizing the metal ions are all conventional ones which can easily be derivatized, show no or only slight nonspecific adsorption, show good physical, mechanical and chemical stability, and have a high external and internal surface area. Suitable materials can be obtained commercially, for example, from Pharmacia LKB, Sweden (Sepharose™6B or Superose™), Pierce, USA (immobilized iminodiacetic acid I or II, immobilized tris(carboxymethyl)-ethylenediamine), Sigma, USA (immobilized iminodiacetic acid-agarose), Boehringer Mannheim, Germany (zinc chelate-agarose) or Toyo Soda, Japan (TSKgel Chelate-5PW). EP-B-0 253 303 describes further suitable materials. Further suitable and advantageous materials are materials such as Ni-coated microtiter plates (nickel-chelate coated Flashplate®, NEN life science products) or magnetic particles or specifically metal ion-treated and binding membranes.

The various metal ions are bound in suitable materials advantageously via groups such as IDA (= iminodiacetic acid), NTA (= nitrilotriacetic acid) or TED (= tris(carboxymethyl)-ethylenediamine). Suitable metal ions are Co, Cu, Fe, Ca, Mg, Ni, Al, Cd, Hg or Zn, preferably Fe, Ni or Cu, particularly preferably Ni or Cu, very particularly preferably Ni. The loading

of the materials with metal ions advantageously takes place with 0.1 to 0.4 M solutions of the metal salts in aqueous, unbuffered solution.

After the washing, the fusion protein is eluted with a suitable buffer. This buffer abolishes the affinity linkage between the fusion protein and the immobilized metal ions. The fusion proteins can be eluted via a pH gradient (low pH values < pH 7.0 act to elute), competitive ligands such as imidazole, organic solvents such as acetone or ethanol, chelating agents such as EDTA or NTA and/or detergents such as Tween 80. Elution via competitive ligands such as imidazole and/or detergents is preferred. Imidazole is used for the elution in a range from 0.05 to 0.7 M, preferably from 0.1 to 0.5 M. The competitive ligands and/or detergents are advantageously employed in a buffer, but can also be used in water. Advantageous buffers are buffers which correspond to the buffers used to load onto the immobilized metal ions. This has the advantage that no unwanted interactions between the column material, the bound proteins and the buffer occur. Advantageous buffers preferably have a pH greater than pH 7, advantageously a pH between pH 7.0 and 9.0, preferably between pH 7.5 and 8.0. These buffers are preferably applied via an increasing gradient. In the case where a pH gradient is used for the elution it is possible to use buffers with a pH below pH 7.0 and/or acids. The eluted fusion protein is collected and can be used immediately or else, if necessary and if desired, treated further. Suitable loading and elution buffers are to be found, for example, in the textbook Protein Purification (Eds. J.C. Janson, L. Rydén, VCH Publisher Inc., 1989, pages 227 to 251).

The protein fragment according to the invention can be deleted using the methods described above, such as cyanogen bromide or protease cleavage. It is possible in this case for residues of the protein fragment to remain in the molecule or else to be completely eliminated from the protein to be purified. The protein fragment is advantageously removed completely from the protein.

Suitable and advantageous protein fragments for preparing fusion proteins can be screened by the following process according to the invention. The invention relates to a process for preparing protein fragments able to enter into a reversible affinity linkage with immobilized metal ions, which comprises carrying out the following steps:

Sub B4

a) preparing a nucleic acid library starting from any suitable nucleic acid sequence which codes for a protein fragment of the sequence

His-X^1-His-X^2-X^3-X^4-Cys-X^5-X^6-Cys,

where the histidine and cysteine residues of the sequence are conserved in the nucleic acid library,

b) fusing the nucleic acids of the library to a reporter gene which makes it possible to detect the fusion protein encoded by the resulting nucleic acid via its binding to the immobilized metal ions and

c) selecting the nucleic acid sequences which display a reversible binding to the immobilized metal ions which is at least 1.5 times stronger than the sequence in the natural Helicobacter pylori ATPase-439.

The nucleic acid library based on the abovementioned sequence can be constructed by methods for mutagenesis known to the skilled worker. For this purpose, the sequence can be subjected, for example, to a site directed mutagenesis as described in D.M. Glover et al., DNA Cloning Vol.1, (1995), IRL Press (ISBN 019-963476-9), Chapter 6, pages 193 et seq.

Spee et al. (Nucleic Acids Research, Vol. 21, No. 3, 1993: 777 - 778) describe a PCR method using dITP for random mutagenesis.

The use of an in vitro recombination technique for molecular evolution is described by Stemmer (Proc. Natl. Acad. Sci. USA, Vol. 91, 1994: 10747 - 10751).

Moore et al. (Nature Biotechnology Vol. 14, 1996: 458 - 467) describe combination of the PCR method and recombination method.

The use of an in vitro recombination technique for molecular evolution is described by Stemmer (Proc. Natl. Acad. Sci. USA, Vol. 91, 1994: 10747 - 10751). It is also possible to use a combination of the two methods.

The use of mutated strains with defects in the DNA repair system is described by Bornscheuer et al. (Strategies, 11, 1998: 16 - 17). Rellos et al. describe a PCR method using non-equimolar amounts of nucleotides (Protein Expression and Purification, 5, 1994: 270 - 277).

The nucleic acid library can advantageously be produced by a PCR technique using two complementary, degenerate oligonucleotides (called wobble primers), as described in the examples. It is important that the histidine and cysteine residues present in this sequence are conserved.

For the screening for protein fragments having an improved metal-binding affinity, the nucleic acid fragment according to the invention, which codes for the protein fragment, is fused to a reporter gene. Advantageous reporter genes make it easy to detect the binding to the immobilized metal ions via, for example, binding of antibodies which are labeled with a fluorescent dye and are directed against the reporter gene, or via self-fluorescing proteins such as the advantageous and preferred egf protein from E.coli (= green fluorescent protein, see Prasher et al., Gene 111 (2), 1992: 229 - 233) or the preferred bioluminescence protein from Aequoria victoria or via light-generating proteins such as the luciferin/luciferase system. Particularly preferably used is a gfp protein mutant (= egfp = enhanced green fluorescence protein) with a 35-fold higher fluorescence activity caused by two point mutations at position 64, replacement of Phe by Leu, and position 65, replacement of Ser by Thr. This protein mutant has the advantage over the wild-type protein that it is soluble and forms no inclusion bodies. The use of the egfp protein makes it possible to locate and quantify the protein concentration in each phase of the purification of the proteins without interfering with the purification and without using other cofactors or substrates (Poppenborg et al., J. Biotechnol., 58 (2), 1997, 77 - 88). The egfp protein is also distinguished by a high stability toward a wide pH range (pH 5.5 to 12), bleaching by photooxidation, oxidizing and weakly reducing agents such as 2% mercaptoethanol. The protein shows a decrease in fluorescence above 37°C. Likewise suitable as reporter gene are the gfp-uv (blue fluorescence) and the eyfp (yellow fluorescence) proteins.

Sub E1

~~The suitable sequences are selected by comparing with the binding affinity to the immobilized metal ions of the following natural Helicobacter pylori ATPase-439 sequence His-Ile-His-Asn-Leu-Asp-Cys-Pro-Asp-Cys. The protein fragment sequences according to the invention show a reversible binding to the immobilized metal ions which is at least 1.5 times stronger, preferably at least twice, and particularly preferably at least three times, stronger. Advantageous sequences make it possible for the protein yield after the purification to be at least 20%, preferably at least 30%, particularly preferably at least 40%, very particularly preferably at least 50%.~~

The process according to the invention for screening the nucleic acid library is advantageously suitable for automation. This process can be used easily for testing a large number of nucleic acid fragments and protein fragments for their metal ion binding affinity in so-called high-throughput screening.

Proteins can easily be detected using the protein fragments according to the invention. In the method according to the invention for detecting proteins, individual proteins which comprise a protein fragment having the abovementioned protein fragments according to the invention in a protein mixture are detected via antibodies which are directed against the protein framework. Detection of these fusion proteins advantageously takes place via mono- or polyclonal antibodies directed against the protein fragment (= tag). The protein mixture can advantageously be fractionated by chromatography or electrophoresis before the detection and subsequently be transferred (= blotted) to a suitable membrane (e.g. PVDF or nitrocellulose) by conventional methods (see Sambrook et al.). This membrane is then incubated with an antibody directed against the tag. It is advantageous to wash the membrane several times and then to detect the bound antibodies via a specific reaction with a second antibody which is, for example, enzyme-conjugated (e.g. alkaline phosphatase, peroxidase etc.) and is directed against the constant region of the first, in a Western blot or immuno blot. Corresponding antibodies are commercially available. Where magnetic particles are used, the washing can be omitted and the antibody-coated magnetic particles can be purified by fishing out with magnets.

The protein fragments according to the invention have the advantage over the conventional His tags for protein detection that they have a stronger antigenic effect and thus are more suitable for producing antibodies against the tag.

The invention is illustrated further by the following examples.

Examples:

The Chelating Sepharose Fast-Flow from Pharmacia LKB, Uppsala, Schweden, was used for the test of binding to metal chelate columns (= immobilized metal ions). Ampicillin, imidazole, EDTA and all other reagents were purchased from Fluka (Buchs, Switzerland). The DNA gel extraction kit, the Midi Plasmid-Kit and the Prepspin Plasmid Kit originated from Qiagen (Hilden, Germany), the restriction enzymes, the DNA-modifying enzymes, the T4 DNA ligase and the Taq polymerase originated from MBI

Fermentas (St. Leon-Rot, Germany). The Taq Dye Cycle Sequencing Kit was purchased from Applied Biosystems (Weiterstadt, Germany).

The E. coli strain DH5α (F^- endA1 hsdR17 [rk^-, mk^+] supE44 thil λgyrA96 relA1 Δ (argF laczya) U169) was used for the cloning experiments. The plasmids which contained the gene for the egf protein were purchased from Clonetech USA. The E. coli strains were cultured in Luria-Bertani medium (= LB) with 100 µg/ml Ampicillin at 37 °C to select the clones for transformation with the egfp vector. The lysis buffer contained 50 mM sodium phosphate buffer pH 8.0, 300 mM NaCl, 1 mg/ml Lysozym and 1 mM PMSF (= phenylmethanesulfonyl fluoride = specific trypsin and chymotrypsin inhibitor).

The DNA methods, such as ligations, restrictions, PCR or transformations etc., were carried out as described in Sambrook, J. et al. (1989) Molecular cloning: A laboratory manual, Cold Spring Harbor Laboratory Press or F.M. Ausubel et al. (1994) Current protocols in molecular biology, John Wiley and Sons. The fluorescence-labeled dideoxy-DNA sequencing method was used for the sequencing. The DNA sequencing was carried out using the Taq Dye Deoxy™ Cycle Sequencing Kit (Applied Biosystems) and the 373A DNA Sequencing System (Applied Biosystems) in accordance with the manufacturer's instructions.

Example 1: Preparation of randomly mutagenized N-terminal metal-binding sites which were bound to the egfp gene, and of his6-egfp

For the PCR, the plasmid egfp and the two following complementary oligonucleotides

5'-GCAATACCATGGGGCATNNNCATNNNNNNNNNTGTNNNNNNTGTGTGAGGAAGGGCGAG-3'

5'-CAGTTGGAATTCTAGAG-3'

were used. In the case of his6-egfp, the following two complementary primers

5'-GCAATACCATGGGGCATCATCATCATCATCATGTGAGGAAGGGCGAG-3'

5'-CAGTTGGAATTCTAGAG-3'

were used.

The conditions used for the PCR were as follows:

Mixture:	8	µl of dNTP mix (200 µmol)
	10	µl of 10 x ThermoPol buffer (New England Biolabs)
	1	µl (100 pmol) of primer 1
	1	µl (100 pmol) of primer 2
	1	µl (100 ng) of egfp plasmid
	79.5	µl of water
	1	µl of Deep Vent polymerase (New England Biolabs)

PCR program	95°C 7min
	95°C 1min
	56°C 1min 30x
	72°C 3min
	72°C 7min

The PCR products were each digested with NcoI and NotI and ligated into the egfp vector which had been digested with the same enzymes, in order to preclude mutations in the vector (see Figure 1). The PCR-vector ligations were used to transform E. coli. The transformants were plated out on LB agar with 100 µg/ml Ampicillin and incubated at 37°C.

Example 2: Cultivation conditions and preparation of the cell lysates

Transformed colonies which showed fluorescence, and some which showed no fluorescence, were selected and cultured in 50 ml of LB medium which contained 100 µg/ml Ampicillin. Colonies showing fluorescence were selected for the high-throughput screening and were incubated in sterile microtiter plates which contained 250 µl of LB medium with 100 µg/ml Ampicillin. After incubation, the cultures were centrifuged. The pellets were resuspended in 2 ml of lysis buffer, incubated on ice for 20 minutes and then disrupted with ultrasound (twice, 5 minutes with a Branson Sonifier 250). After centrifugation (15 min, 4°C, 20000 x g) the various egfp mutants were obtained in the supernatant. All the selected clones were sequenced (see Tables I and II). Clones which showed no fluorescence contained stop codons in the sequence, so that no functional proteins were expressed (see Table I, A8, A13, M16a, Z4, Z11 and Z13). To quantify the bound proteins, a fluorescence measurement was carried out in all the experiments and correlated in a wide range with the gfp concentration (see Figure 2).

Example 3: Low-throughput screening with Ni-NTA columns from Qiagen

600 µl of the lysed cells were loaded onto an Ni-NTA column, washed twice with 600 µl of a washing solution (50 mM sodium phosphate buffer, pH 8.0, 250 mM NaCl) and then eluted with a 0.7 M imidazole solution.

Example 4: High-throughput screening with membrane filter plates

Membrane filter microtiter plates (MultiScreen 5 µm; supplied by Millipore, Molsheim, Germany) were used for the high-throughput screening. 250 µl of a stirred chelate Sepharose suspension were placed in each well of the membrane filter plates three times. After each addition, the Sepharose was centrifuged down (2 min, 23°C, 350 rpm). All further steps were carried out in a Beckmann Biomek 2000 robot. The minicolumns in the wells were washed twice with 250 µl of water. The Sepharose was then loaded with 250 µl of a metal salt solution and equilibrated three times with 200 µl of buffer (50 mM sodium phosphate, pH 8.0, 250 mM NaCl). 0.3 M $NiCl_2$, 0.3 M $CuSO_4$ or 0.3 M $ZnCl_2$ solutions were used for each of the various mixtures (loading with metal ions). Aqueous unbuffered metal salt solutions were used. 250 µl of the cell lysate supernatants were placed on each of these minicolumns. The columns were then washed twice with 250 µl of equilibration buffer. The bound proteins were eluted with 2 × 100 µl of 0.5 M imidazole in equilibration buffer. The chelate Sepharose in the filters can be regenerated with 250 µl of 50 mM EDTA, 1 M NaCl in water and be used for further screening experiments.

Example 5: IMAC experiments

A conventional chromatography system consisting of a glass column, two peristaltic pumps for applying the solutions, a UV detector (LKB UV-MII), a printer (LKB RIC 102) and a fraction collector (LKB FRAC-200) was chosen for the experiment. All the apparatus originated from Pharmacia. The column was packed with chelating Sepharose Fast-Flow Gel (Pharmacia), washed with 7 bed volumes of deionized water and loaded with the metal ions with 7 bed volumes of 0.3 M $NiCl_2$ solution. The column was then washed and equilibrated with 7 bed volumes of IMAC buffer (50 mM sodium phosphate, pH 8.0, 250 mM NaCl). 1 ml samples of the cell lysates were loaded onto the column at a flow rate of 1.5 ml/min and washed with 10 bed volumes of IMAC buffer. The bound proteins were eluted via an increasing gradient with 0.5% of a 0.5 M imidazole solution per ml of elution solution and finally 5 bed volumes of a 0.5 M imidazole/water solution. The protein fractions were identified by UV detection and collected. After completion, the column was washed and regenerated with 50 mM EDTA/1 M NaCl solution. This washing step detaches the metal, bound cell residues and proteins from the column. The eluted

fractions were examined both optically and spectroscopically. Some of the investigated clones showed no affinity for the matrix (see Tab. I, M15, M16), while others showed good binding (see Tab. I, M13, Z5 and Z7). The clones M13 and Z5 eluted from the column in a sharp band, whereas there was smearing of clone Z7 on the column.

Example 6: Experiment comparing egfp wild-type and his-tag

The clone M13 was compared with the egfp wild-type protein and the usual his tags in a comparative experiment. The egfp wild-type protein does not bind to the metal chelate columns. Fluorescence was no longer detectable on the column after washing the column. The clone M13 binds to the column in a sharp band, whereas the his tag proteins bind over the entire column. This is attributable to a lower affinity, which leads eventually to a lower capacity of the column. The protein yield in the case of M13 is 56%, which is higher than the 48% with the his tags.

Table I: Binding experiments with Ni metal chelate columns

Sub E2

Clone	Amino acid sequence									
A6	His	Gln	His	Glu	Gly	Arg	Cys	Lys	Glu	Cys	gfp
A8	His	Cys	His	Pro	Glu	Leu	Cys	Stop	Leu	Cys	gfp
A13	His	Leu	His	Ser	Ile	Gly	Cys	Pro	Stop	Cys	gfp
M13	His	Asn	His	Arg	Tyr	Gly	Cys	Gly	Cys	Cys	gfp
M14	His	Ser	His	Ser	Val	Gly	Cys	Phe	Phe	Cys	gfp
M15	His	Gly	His	Thr	Leu	Ser	Cys	Gly	Leu	Cys	gfp
M16	His	Ser	His	Thr	Leu	Arg	Cys	Lys	Gly	Cys	gfp
M16a	His	Ser	His	Stop	Leu	Arg	Cys	Lys	Gly	Cys	gfp
Z4	His	Stop	His	Asn	Stop	Val	Cys	Ala	Thr	Cys	gfp
Z5	His	Arg	His	Gly	Thr	Asn	Cys	Leu	Lys	Cys	gfp
Z7	His	Ile	His	Gln	Ser	Asn	Cys	Gln	Val	Cys	gfp
Z11	His	Thr	His	Ala	Ser	Gly	Cys	Stop	Stop	Cys	gfp
Z13	His	Cys	His	Thr	Trp	Cys	Cys	Asn	Stop	Cys	gfp

The clones A6, A10, M13, Z5 and Z7 bind well to the metal chelate column, whereas the clones M14, M15 and M16 showed no binding.

Table II: Further sequenced clones detected by fluorescence in the High-throughput screening

Sub E3

Clone	Amino acid sequence										
A1	His	Gly	His	Met	Glu	Arg	Cys	Leu	Val	Cys	gfp
A2	His	Lys	His	Ala	Arg	Ser	Cys	Met	Gly	Cys	gfp
A3	His	Phe	His	Thr	Val	Phe	Cys	Phe	Ser	Cys	gfp
A4	His	Arg	His	Arg	Gly	Met	Cys	Thr	Ala	Cys	gfp
A12	His	Asp	His	Arg	Gly	Val	Cys	Gly	Leu	Cys	gfp
A14	His	Asp	His	Glu	Arg	Leu	Cys	His	Asn	Cys	gfp
X8	His	Gly	His	Gly	Asn	Arg	Cys	Cys	Gly	Cys	gfp
X9	His	Arg	His	Gly	Thr	Ala	Cys	Met	Asp	Cys	gfp
X11	His	Ile	His	Ile	Met	Thr	Cys	Leu	Ser	Cys	gfp
X12	His	Thr	His	Pro	Arg	Ser	Cys	Ala	Glu	Cys	gfp
X15	His	Gly	His	Asp	Arg	Thr	Cys	Arg	Gly	Cys	gfp
X16	His	Arg	His	Ala	Ile	Ser	Cys	Ile	Gly	Cys	gfp
X17	His	Ile	His	Arg	Gly	Asp	Cys	Tyr	Glu	Cys	gfp
X18	His	His	His	Gly	Ser	Thr	Cys	Pro	Thr	Cys	gfp
X19	His	His	His	Phe	His	Ser	Cys	Phe	Tyr	Cys	gfp
Z8	His	Lys	His	Val	Asp	His	Cys	Gly	Arg	Cys	gfp
Z9	His	Ser	His	Leu	Thr	Leu	Cys	Leu	Gly	Cys	gfp
Z10	His	Thr	His	Gln	Ser	Gln	Cys	Gly	Arg	Cys	gfp
Z14	His	Arg	His	Leu	Phe	Trp	Cys	Ser	Glu	Cys	gfp

Example 7: Metal binding affinity

Many of the clones showed a preferred binding to Ni^{2+} or Cu^{2+}. In the case of M13, no binding to Zn^{2+} was observed. On use of Ni chelate columns, the clone M13 showed distinctly better purification of the proteins by comparison with the his tags. Conversely, the latter resulted in a purer product by comparison with M13 on use of Cu chelate columns, but, since the binding to the column material is very strong in both cases, Cu ions were washed out by the drastic elution conditions. This leads to contamination of the products.

Example 8: Experiment comparing between the ATPase-439 sequence and protein fragments according to the invention

The ATPase-439 comparison clone was carried out in analogy to Example 1 and 6. The primer used was the following primer 5'-GCAATACCATGGGGCATATTCATAATCTTGATTGTCCTGATTGT-3'. The other primers and the PCR conditions were as described in Example 1.

The experiment was carried out with a Qiagen Ni-NTA Spin Kit under native conditions, lysis being carried out as described under these conditions; the columns which had already been loaded were equilibrated with 600 µl of 50 mM sodium phosphate buffer, pH 8.0, 300 mM NaCl and centrifuged at 2000 rpm (= 420 × g) for 2 min. Then 600 µl of the lysate were applied and centrifuged for 2 min. Two washes were carried out with 600 µl of a 50 mM sodium phosphate buffer, pH 8.0, 300 mM NaCl each time. Then 600 µl of 50 mM sodium phosphate buffer, pH 8.0, 300 mM NaCl, 0.5 M imidazole were used for two elutions. The yield of pure protein was 1.5 times higher with the clone M13 than the metal-binding site of the ATPase-439 comparison clone. Clone M13 thus binds better and can also be eluted better.